



## 저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

공학석사 학위논문

# 스마트폰으로 촬영된 알약 영상의 글자 및 형상 인식 방법

Shape and Text Imprint Recognition of Pill  
Image Taken with a Smartphone

2017년 8월

서울대학교 대학원

기계항공공학부

김 대 욱

## 초 록

삶을 이롭게 하기 위해 수 많은 약들이 제조 · 판매되지만 이러한 약들은 오용되거나 남용될 경우 사람에게 치명적인 문제를 일으킬 수 있다. 인터넷과 스마트폰을 통해 알약을 검색하고 그에 대한 정보를 얻을 수 있으나, 모양과 색상, 글자를 직접 입력 해야하고 특수한 마커를 배경으로 사용해야하는 등 접근성이 낮다. 따라서 특수한 마커를 사용하지 않으면서도 알약 이미지로부터 글자 정보까지 획득할 필요가 있다.

본 논문에서는 스마트폰으로 촬영된 알약 영상에서 글자와 형상을 인식하는 방법을 제안한다. 알약이 포함된 영상에서 알약 영역을 특정 짓기 위해 Saliency Map을 이용한 뒤, 빛 효과와 그림자 효과를 제거한다. 이렇게 획득된 알약 영역에서 Zernike Moment를 통해 형상 정보를 얻는다. Gaussian Filter, Gradient Filter, Binarization을 통해 알약의 글자를 감싸는 박스를 추출하고, CNN Deep Learning을 통해 학습된 각인 글자 인식기를 사용하여 최종적인 글자 정보를 획득한다. 본 논문에서 제안하는 알고리즘을 테스트하기 위해 NLM 데이터베이스의 알약 이미지 총 500개를 사용하였다. 또한, 1) Shape Matching Rate, 2) Text Box Detection Rate, 3) Character Recognition Rate, 4) Text Recognition Rate, 5) Recognition Success Rate로 총 5가지 항목에 대해 알고리즘을 평가하였다. 그 결과는 각각 75.5%, 87.5%, 0.786,

73%, 58.4%였으며, 기존의 알약 인식 알고리즘과 비교하였을 때, 상당히 개선된 결과를 얻었다.

**주요어** : 글자 인식, 딥 러닝, 알약 인식, 영상 검색, 영상 처리, Zernike 모멘트

**학 번** : 2014-21871

# 목 차

초록 .....	i
목차 .....	iii
표 목차 .....	v
그림 목차 .....	vi
제 1 장 서론 .....	1
제 2 장 관련연구 .....	4
2.1 모양 또는 색상 등 일부만 인식 .....	4
2.2 모양과 색상, 글자 모두 인식 .....	5
제 3 장 제안된 인식 시스템 .....	10
3.1 전체 시스템 개요 .....	10
3.2 알약 주변부 획득 .....	11
3.3 알약 영역 추출 .....	13
3.4 형상 정보 획득 .....	14
3.5 글자부 획득 .....	19
3.6 글자 학습 및 인식 .....	22
제 4 장 실험 및 결과 .....	25
4.1 Database .....	25
4.2 Shape Matching Rate .....	26

4.3 Text Box Detection Rate .....	29
4.4 Character Recognition Rate .....	30
4.5 Text Recognition Rate & Recognition Success Rate.....	32
제 5 장 결론 .....	34
참고문헌.....	35
Abstract .....	39

## 표 목차

Table 1. Text Box Detection Rate Result.....	29
Table 2. Character Recognition Rate Result.....	31
Table 3. Text Recognition Rate & Recognitnion Success Rate Result .....	32

## 그림 목차

Figure 1. 알약 인식을 위해 사용된 마커 .....	4
Figure 2. Jiye Yu et al.에서 사용한 알약 이미지 .....	6
Figure 3. 실제로 사용자가 촬영한 알약 영상 .....	7
Figure 4. 촬영 환경에 따라 다양하게 나타나는 알약의 색 .....	9
Figure 5. 전체 시스템 개략도 .....	10
Figure 6. 알약 영상과 Saliency Map .....	12
Figure 7. 그림자 제거 과정 .....	14
Figure 8. LAB 색상 공간에서 표현된 알약 영상 .....	16
Figure 9. 알약 영역 추출 예시 이미지 .....	17
Figure 10. 분류에 사용된 14가지 모양 클래스 .....	18
Figure 11. 글자부 강화 과정 .....	19
Figure 12. CNN Deep Learning의 원리 .....	22
Figure 13. 각인 글자 인식을 위해 사용된 CNN 네트워크 구성 .....	23
Figure 14. 이미지 데이터베이스 (NLM) .....	25
Figure 15. Rank에 따른 모양 인식률 .....	26
Figure 16. Rank에 따른 모양 인식률 (Reference 이미지를 이용한 비교) .....	28
Figure 17. Text Box Detection Rate 계산 방식 .....	29



# 제 1 장 서 론

사람들의 삶을 이롭게 하기 위해 수 많은 약들이 만들어진다. 그러나 이러한 약들은 오용되거나 남용될 경우 사람에게 치명적인 문제를 일으킬 수 있다. 특히 현대에 이르러서는 약의 종류가 다양해졌기 때문에 단순히 크기와 모양, 색깔 등으로만 판단하는 것은 힘들고, 약에 적힌 글자까지 읽어야 약을 구분할 수 있게 되는 경우가 많아졌다. 약을 자주 복용하게 되는 노인들은 알약에 작은 글씨로 쓰여 있는 글자를 읽는 것이 쉽지 않기 때문에 약을 구분하는 능력이 일반인들보다 더 떨어진다. 따라서 알약의 글자를 읽어내야 할 필요성은 더 중요해진다.

알약의 오용 및 남용 문제를 줄이기 위하여 다양한 웹사이트에서는 알약 검색 기능을 제공하고 있다. 대표적으로 한국에서는 Naver[1]와 약학정보원[2]이 있으며, 외국에서는 WebMD Pill Identification Tool[3], RxList[4], Healthline[5], DailyMed[6], Drug.com[7] 등이 있다. 이러한 웹사이트들에서 제공하는 검색 기능은 2가지로 나누어져 있는데, 첫번째는 단순히 약의 이름과 성분으로 검색하는 방법이고 두번째는 약의 외형을 통해 검색하는 방법이다. 첫번째 방법은 약의 이름이나 성분을 알고 있어야 한다. 이는 일반인들이 사용하기에는 부적합하다. 즉, 두번째 방법이 조금 더 일반인들이 검색하기에 쉬운 방법이라고 할 수 있다. 그러나 두번째 방법을

사용하기 위해서는 모양, 그리고 색상, 약에 새겨진 글자, 여기에 추가적으로 분할선 유무나 제형 등을 입력해야한다. 특히 약에 무늬가 새겨진 경우에는 입력이 불가능하며, 알약의 색상 또한 주변 빛에 따라 얼마든지 변할 수 있으며, 글자 입력이라는 방식은 시력이 좋지 못한 노인들이 쉽게 사용 가능할 것인가를 생각해보았을 때, 그 실용성이 많이 낮다고 볼 수 있다.

본 논문에서는 일반인도 쉽게 이미지를 통해 알약을 검색할 수 있는 시스템을 제안한다. NLM(National Library of Medicine)[8]에서 구축한 알약 영상 데이터를 이용하여 알약의 모양과 알약에 새겨진 글자를 인식한다. 먼저 알약이 포함된 영상에서 알약 영역을 특정 짓기 위해 Saliency Map[9]을 만든다. 그 뒤, 알약과 알약 주변부에 존재하는 빛 효과 및 그림자 효과를 제거한다. 이렇게 빛 효과가 제거된 영상에서 Canny Edge[10]와 LAB Color Space를 이용해 알약 영역을 형상과 함께 추출한다. 여기서 얻어지는 알약 형상 영상은 이진 영상이기 때문에 Zernike Moment[11] 값을 계산하고 형상 정보를 획득한다. 알약 영역은 글자 인식을 위해 세로 길이를 일정한 값으로 정규화한다. Gaussian Filter를 통해 노이즈를 제거하고, 이미지의 밝기를 조정해 다음 Gradient Filter를 통해 각인 글씨를 뚜렷하게 하고 글자 영역을 추출하게 된다. 마지막으로 CNN Deep Learning을 통해 학습시킨 각인 글자 이미지를 인식시킴으로써 글자를 최종적으로 획득한다.

본 논문의 순서는 다음과 같다. 먼저 2장에서는 기존에

제시되었던 연구들과 그 한계들을 살펴보고 3장에서는 알약의 형상과 글자를 추출하는 알고리즘을 소개한다. 제시하는 인식 시스템을 평가하기 위한 실험과 결과에 대한 내용은 4장에서 서술하며 5장에서는 결론을 서술한다.

## 제 2 장 관련 연구

### 2.1 모양 또는 색상 등 일부만 인식

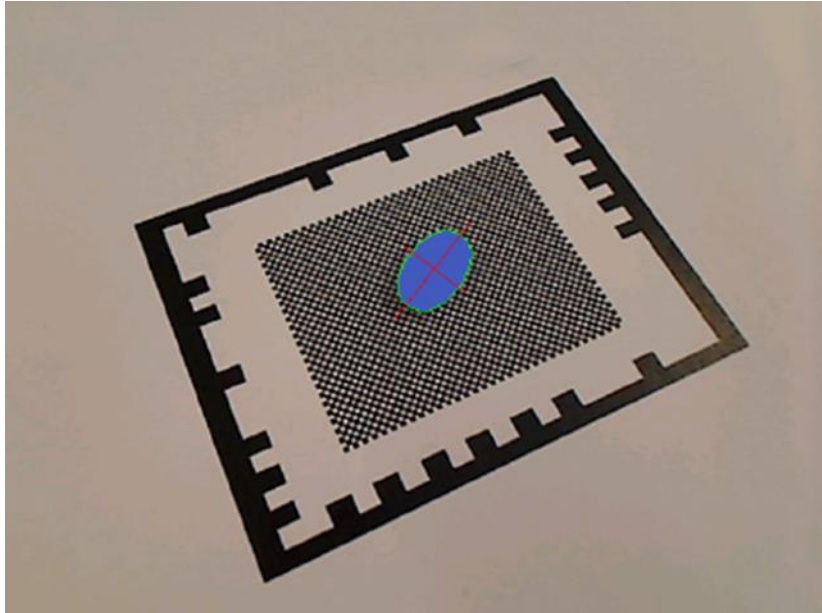


Figure 1. 알약 인식을 위해 사용된 마커[12]

알약의 모양이나 색상만, 혹은 둘 다를 인식하고자 하는 연구들은 진행된 바 있다. Hartl et al.[12]은 Figure. 1과 같이 특수한 마커가 인쇄된 종이를 사용하여 이 마커 위에 알약을 올려두고 스마트폰으로 동영상 촬영한다. 그 뒤, Feature Matching을 통하여 3D Reconstruction 함으로써 실제 크기와 형상, 색깔 정보를 획득했다. 이 방법은 동영상의 화질과 3차원 재건 Resolution의 한계가 있기 때문에 글자 정보를 전혀 얻을 수가 없다는 단점이 있다. 비슷한 방법으로 Cunha et al.[13] 또한 흑색 바탕의 마커를 사용하였는데

동영상인 아닌 단순한 사진을 이용했으며, 사용된 알약의 개수가 14개 밖에 되지 않아 다양한 알약 종류에 대응하기 힘들다. 이러한 기존 연구들에서는 마커라는 추가 도구를 사용하기 때문에 마커가 없으면 인식이 불가능하고 이에 따른 불편함이 존재한다.

마커를 사용하지 않고 모양과 색상 정보를 이용한 연구들도 있다. [14]에서는, 색상 정보는 HSV 색상공간을 이용하고 모양 정보는 Canny Edge로 외곽선을 검출한 뒤, 중심에서 외곽선까지의 거리를 정규화하여 비교하였다. 모양 정보를 단순히 중심 정렬 및 크기 정규화를 통해 비교한 연구[15]도 존재한다. 그러나 이러한 방식의 문제점은 모양 정보를 획득함에 있어, 노이즈에 취약하다는 것이다. 따라서 알약의 각도가 틀어진다거나 알약의 형상이 미묘하게 잘못 추출될 경우 완전히 다른 형상으로 인식될 수 있다.

모양과 색상 정보만을 이용하는 경우 검색하고자 하는 알약의 종류가 많아질수록 검색 성능은 떨어질 수 밖에 없고 더 많고 확실한 정보가 필요하게 된다. 이러한 정보가 바로 다음 절에서 설명할 연구들에서 사용한 글자 정보이다.

## 2.2 모양과 색상, 글자 모두 인식

Lee. B.Y. et al.은 처음에 알약에 새겨진 글자에 Sobel Edge를 적용하고, 영상 이진화 및 영상 처리를 통해 다양한 이미지를 만들었다. 이를 Image Descriptor로 사용하고, 이후 Hu Moment[16] 값을 통해

알약을 분류하였다[17]. 이 연구를 더 발전시켜서 모양 구분에 Hu Moment를 사용하고 색상은 히스토그램, 글자는 SIFT Descriptor[18]와 MLBP Descriptor[19]를 사용하여 알약을 분류하기도 하였다[20]. 그러나 글자를 OCR과 같은 방법으로 읽은 것이 아니라 Descriptor를 사용했기 때문에 빛에 의한 노이즈나 각인된 글자가 조금 뭉개지는 경우 인식률을 더 개선시킬 수 없다는 한계가 있다.

영상 이진화를 통해 글자를 어느 정도 추출한 다음 Descriptor를 통해 인식률을 높인 연구도 있다[21,22]. 이 연구에서는 자체적으로 개발한 기기를 이용해 알약 데이터를 수집하였으며 따라서, 주변 빛에 의한 노이즈나 촬영 조건 등이 상당히 제한되어 있다.

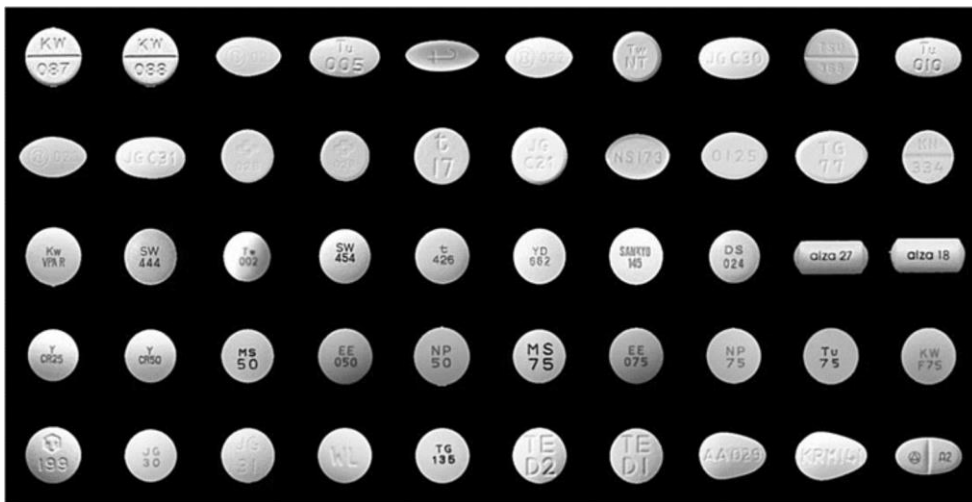


Figure 2. Jiye Yu et al.[21,22]에서 사용한 알약 이미지



Figure 3. 실제로 사용자가 촬영한 알약 영상

Figure. 2는 해당 논문에서 사용한 알약 이미지이다. 잉크로 새겨진 알약의 비중이 50% 정도이고 나머지 각인 알약도 글자 부분은 상대적으로 어두운 형태를 띄고 있기 때문에 단순히 영상을 이진화하기만 해도 글자를 뽑아낼 수 있다. 비교를 위해 Figure. 3의 실제로 사용자가 촬영한 알약 그림을 보면, 글자 영역의 밝기가 주변과 크게 다르지 않고 글자 영역 내에서도 밝기 변화가 복잡하게 나타나는 것을 알 수 있다.

마찬가지로 Suntronsuk et al.은 NLM 데이터베이스 중 촬영 조건이 제한된 이미지들을 Canny Edge와 이진화를 통해 글자를 인식하려 했으나 이진화의 기준이 되는 threshold 값을 정하는 것이 어려워 인식률이 좋지 않은 결과를 보였다[23,24].

즉, 지금까지의 연구들을 종합해보면, 모양과 색상 일부만 사용하는 알고리즘들은 광범위한 알약 종류들에 대해서는 적용하기 힘들며, 글자 부의 인식을 성공한 연구들도 제한적인 환경에서 촬영된 알약 영상들만 사용하였다. 따라서 실제 환경에서 촬영된 알약 영상들에서는 그 성능이 떨어질 수 밖에 없고 이에 대한 새로운 해결책이 필요하다.

각 방법의 우열성을 평가한다면, 글자 자체를 인식시키는 방법과 기술자를 이용해 인식하는 방법 사이에서는 글자 자체를 인식시키는 방법이 더 좋다고 볼 수 있다. 이는 일부가 잘못 찾아졌을 경우 연관성을 바탕으로 보완이 가능하고, 이미지가 존재하지 않더라도 검색 데이터베이스를 업데이트 할 수 있기 때문이다. 모양 인식의 경우에는 경계 거리장 방식과 이미지 모멘트 방식이 있는데, 이 둘 가운데에서는 회전과 크기 불변성의 특성을 가지고 있는 이미지 모멘트 방식이 더 우수하다고 볼 수 있다. 마지막으로 색상 정보를 사용하는 것에 대해서, 제한적인 환경에서는 사용이 가능하겠지만 실제 환경에서는 오히려 색상을 통해 분류해버리면 잘못된 결과를 초래할 수 있다. 실제로 Figure. 4에서는 주변광에 의해 알약의 색깔이 다양하게 변화하는 것을 확인할 수 있다. 특히 글자 정보만이라도 정확하게 찾아진다면 모양정보나 색상정보가 없더라도 알약 후보군을 1~2개로 줄일 수 있기 때문에, 색상 정보의 영향력은 크지 않다.





Figure 4. 촬영 환경에 따라 다양하게 나타나는 알약의 색

## 제 3 장 제안된 인식 시스템

### 3.1 전체 시스템 개요

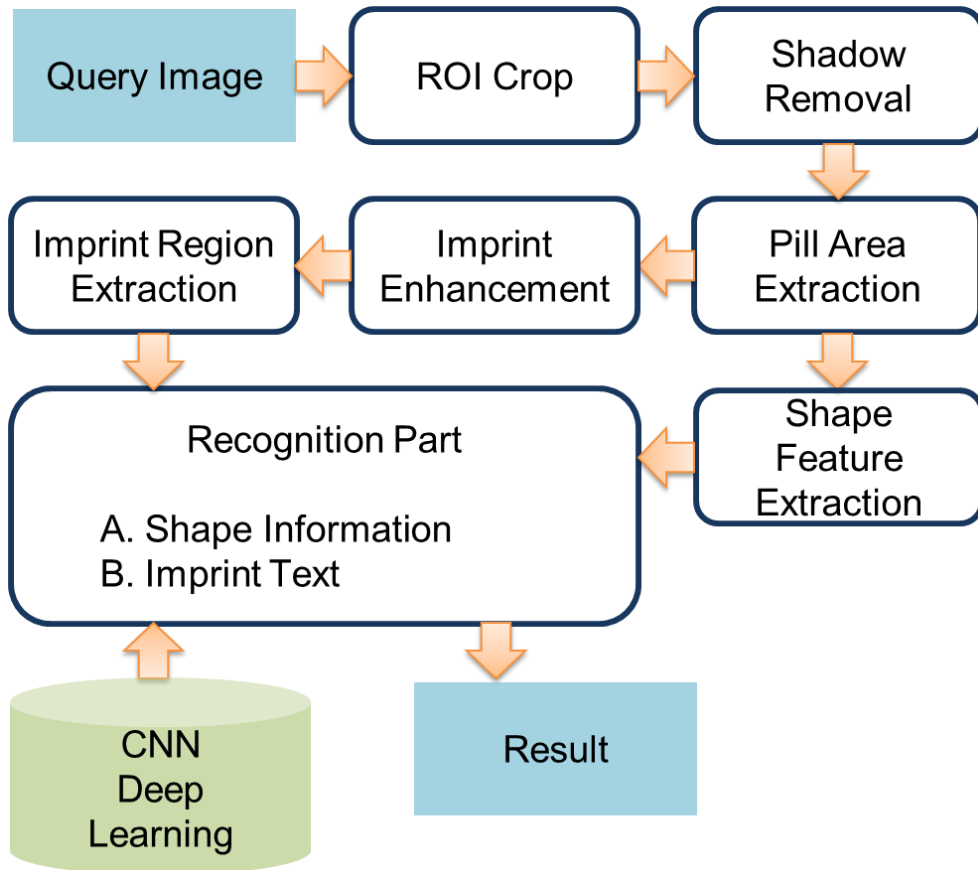


Figure 5. 전체 시스템 개략도

본 논문에서는 앞서 진행되었던 연구들의 문제점을 해결하는 새로운 시스템을 제안한다. 전체 시스템의 개요는 Figure. 5와 같으며, 제안하는 시스템은 여러 방식의 우열성을 따져본 결과를 바탕으로 색상은 제외하고 모양과 글자 정보만을 추출한다. 먼저, 알약 영상은

일반적으로 스마트폰을 사용할 경우, 3000x4000정도의 크기로 저장되기 때문에 이 상태로 영상 처리를 진행하는 경우 전체 시스템의 속도가 느려질 수 있다. 따라서 알약이 촬영된 배경과 알약을 대략적으로 구분하고 알약 주변부를 획득한다. 그 뒤 알약 영역을 정확하게 추출하기 위해서 그림자 및 주변 빛 효과를 제거한다. 배경과의 색상 차이 및 Canny Edge를 이용하여 알약 영역을 추출하고 Zernike Moment를 이용해 형상 정보를 획득한다. 알약 영역은 다시 Gradient filter를 이용해 글자부를 강조하고 글자부의 Bounding Box를 획득한다. 글자 인식은 Deep Learning 학습을 기반으로 앞서 잘라진 각각의 글자부를 읽음으로써 글자 정보를 획득한다. 최종적으로 모양 정보와 글자 정보를 종합하여 알약을 인식한다.

### 3.2 알약 주변부 획득

최근의 스마트폰은 가로 4000 픽셀, 세로 3000 픽셀 수준의 고화질 영상을 저장할 수 있다. 따라서 최근의 스마트폰으로 알약을 촬영한다고 가정할 경우, 영상 또한 고화질로 획득될 것이다. 그러나 이러한 고화질 상태로 영상 처리를 진행하는 경우 전체 시스템의 속도가 저하되기 때문에 영상 처리를 필요로 하는 영역을 뽑아내어 영상의 전체 크기를 줄일 필요가 있다.

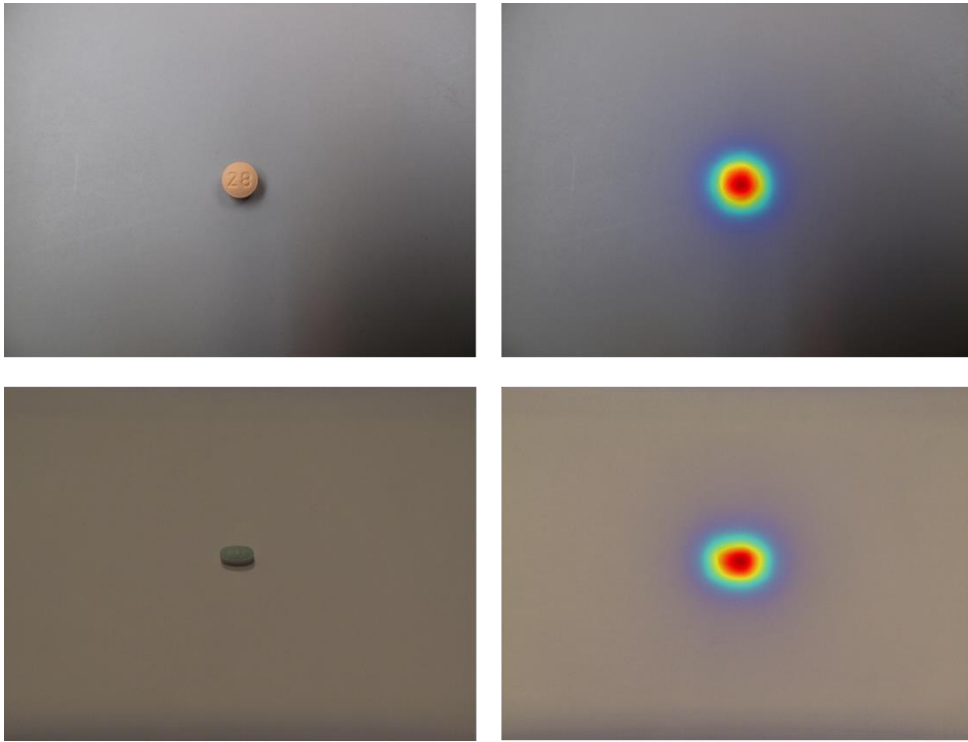


Figure 6. 알약 영상과 Saliency Map

제안된 시스템은 알약과 배경 영역을 구분하기 위해 영상의 사이즈를 10분의 1로 줄인 뒤, Saliency Map[9]을 적용하였다. 알약이 나타나는 영역은 배경과 색상 차이가 있을 뿐만 아니라 알약의 경계 및 그림자 효과에 의해 배경과 뚜렷이 구분된다. Saliency Map은 이렇게 주변과 구분되는 경향을 수치화하여 나타내준다. Figure. 6은 알약 영상에서 Saliency Map을 표현한 것이다. Saliency Map에서 Otsu Threshold를 적용하여 영상을 이진화하면 알약 주변부 영역을 획득할 수 있다. 이때, 노이즈가 섞일 수 있고, 일부 알약이 잘릴 수 있기 때문에 이진 영역에서 가장 큰 영역을 알약 영역으로 가정하고, 이를 둘러싸는 박스를 추출한 뒤, 상하좌우로 300픽셀씩 넓혀서 잘라낸다.

### 3.3 알약 영역 추출

최종적으로는 알약 영역만 따로 뽑아내야 하는데, 주변 빛이나 그림자 효과에 의해서 경우에 따라 이러한 분리가 어려워 질 수 있다. 따라서 이러한 효과를 제거할 필요성이 있다. RGB 색상공간으로 표현된 알약을 Figure. 7(a)와 같이 HSV 색상공간으로 변경하면 어두운 공간과 밝은 공간은 V값에 의해 결정된다. V의 최소값을 0, 최대값을 1이라고 할 때, V가 0.9 이상인 영역은 빛이 과한 영역, V가 0.3 미만인 영역을 그림자 영역, V가 0.3이상 0.5 이하인 영역은 준-그림자 영역으로 정하면 Figure. 7(b)과 같은 이진 영상을 획득할 수 있다. 그 뒤, Gaussian Smoothing을 하면 Figure. 7(c)처럼 Weight 영상을 만들 수 있다. 여기에 이미지 평균 밝기 값을 이용해 차이만큼 보정해주면 최종적으로 그림자가 제거 된 Figure. 7(d)를 획득할 수 있다.

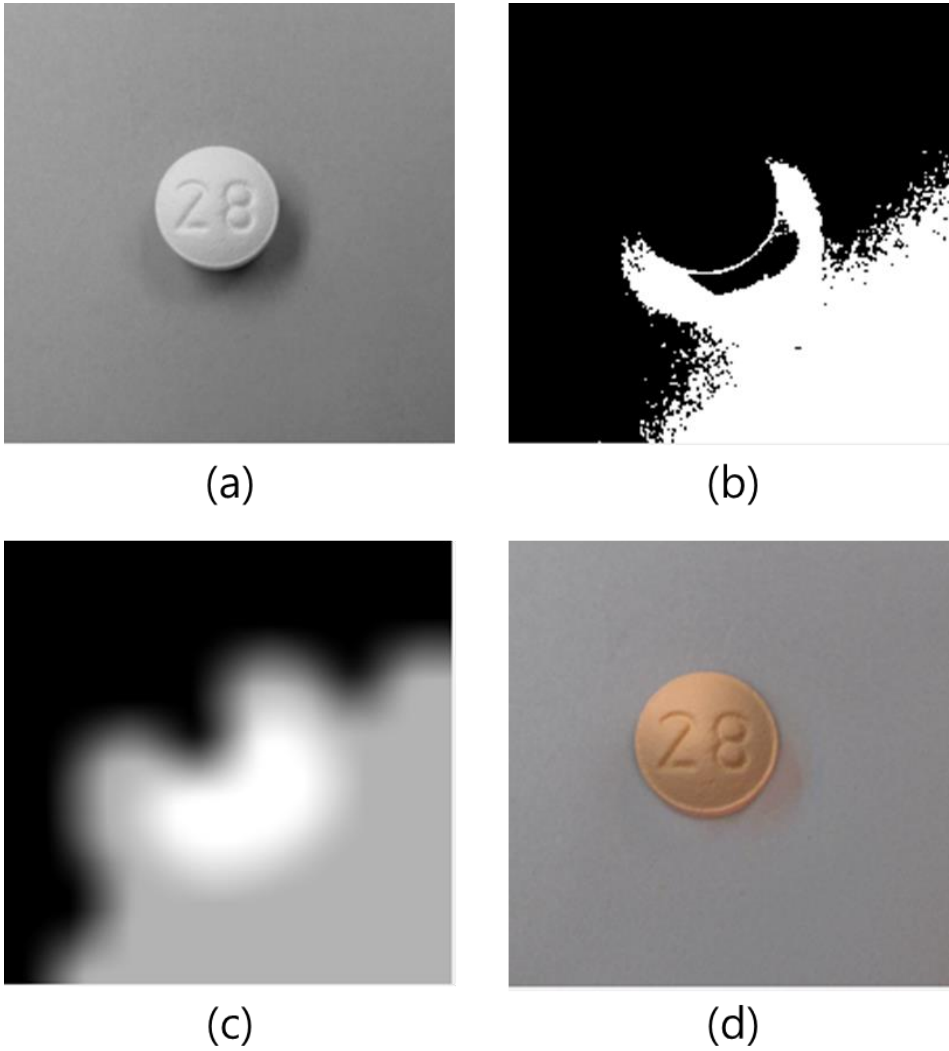


Figure 7. 그림자 제거 과정. (a) HSV 색상 공간 중 V 영상. (b) 준-그림자 영역. (c) Weight 영상. (d) 그림자가 제거된 영상

이를 수식으로 표현하면 다음과 같다. 먼저 각 영역에 해당하는 이진 영상  $B_3$ ,  $B_5$ ,  $\bar{B}_3$ 을 아래와 같이 정의한다.

$$B_3(x, y) = \begin{cases} 1 & \text{if } V(x, y) < 0.3 \\ 0 & \text{otherwise} \end{cases}$$

$$B_5(x, y) = \begin{cases} 1 & \text{if } 0.3 \leq V(x, y) < 0.5 \\ 0 & \text{otherwise} \end{cases}$$

$$\bar{B}_3(x, y) = 1 - B_3(x, y)$$

이 때, 그림자 영역을 제외한 영역의 평균 밝기 값  $m_3$  과 Weighted 영상  $B_w$  는

$$m_3 = \frac{\sum_x \sum_y \bar{B}_3(x, y) \cdot V(x, y)}{\sum_x \sum_y \bar{B}_3(x, y)}$$

$$B_w = B_3 + 0.7B_5$$

와 같다. V영상과 Weighted 영상에  $r=30$ ,  $\sigma=2$  그리고  $r=30$ ,  $\sigma=20$  으로 Gaussian Filter를 각각 적용한 이미지를  $G_v(x, y)$ ,  $G_w(x, y)$  라 하면, 그림자가 보정된 V영상  $\tilde{V}(x, y)$  는

$$\tilde{V}(x, y) = V(x, y) - \{m_3 - G_v(x, y)\} \cdot G_w(x, y)$$

가 된다.

알약은 고유한 색깔을 가지고 있는데, 이 색깔이 주변 배경의 색과 다른 경우도 존재하고, 주변 배경과 색깔이 유사한 경우도 존재한다. 따라서 다양한 경우를 반영하기 위해 총 3가지의 방법으로 알약 영역을 추출한다. 첫 번째는 Canny Edge를 통한 방법이다. 알약의 경계는 배경과의 밝기 차이에서 비롯된다는 사실에 착안한다. 알약의 경계가 완전한 Closed Loop으로 나오면 Loop이 알약의 형상 외곽선이 된다. 실제로는 이렇게 Closed Loop으로 나오는 경우도 있고 경계가 그림자와 뒤 섞여 Open Loop로 나타나는 경우도 존재한다. 이를 보완하기 위해 두번째

방법은 LAB 색상 공간을 이용한다. 배경과 알약의 색깔 차이가 있다면 LAB 색상공간에서 A와 B 둘 중 하나는 값의 차이가 크게 발생한다. 이를 이용하면 알약 형상을 획득할 수 있다. 마지막 방법은 Gradient 영상에 Adaptive Threshold를 적용했다. 이는 밝기도 유사하고 색상도 유사한 경우에, 넓은 범위에서 경계가 될 만한 영역을 가늠하고 이렇게 찾아진 부분 부분의 외곽선을 서로 이어주는 방법이다.

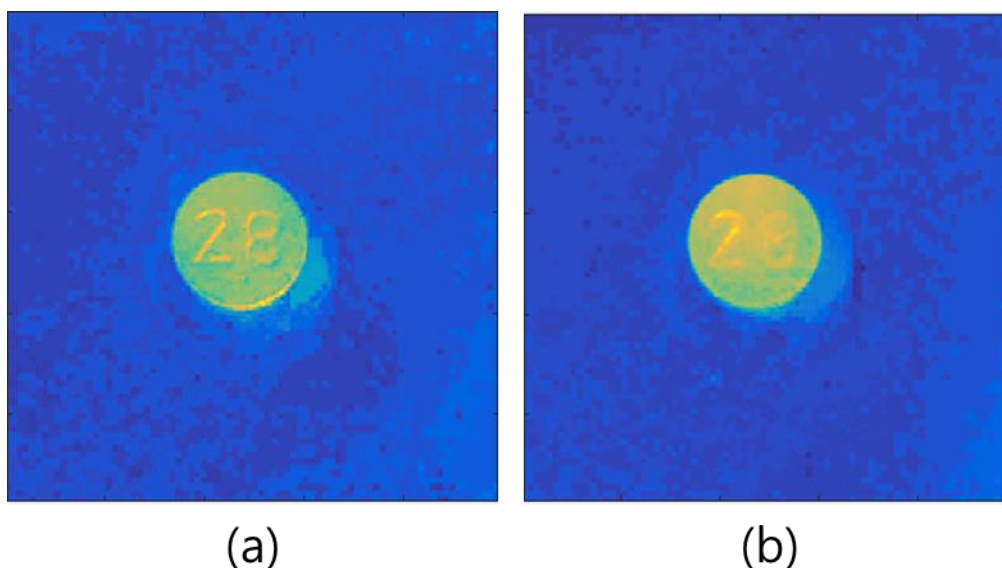


Figure 8. LAB 색상 공간에서 표현된 알약 영상. (a) A 색상공간. (b) B 색상공간

앞서 Figure. 7(d)의 영상에서 LAB 색상 공간으로 변경한 예시가 Figure. 8이다. 색상공간과 B 색상공간은 알약 영역과 배경 영역에서 뚜렷한 차이를 표현할 수 있다는 것을 확인할 수 있다.

이러한 방법을 통해 최종적으로 알약 영역을 추출하게 된다. Figure. 9는 추출된 영역 결과 이미지이다.





Figure 9. 알약 영역 추출 예시 이미지

### 3.4 형상 정보 획득

형상 정보는 알약이 어떤 모양인가를 나타낸다. 실제 Naver의 알약 검색 기능[1]에서 제공하는 모양 및 전준철 et al.의 연구에서 사용된 모양[14]을 고려해 총 14가지의 모양을 정의하였다. 정의한 모양은 Figure. 10과 같다. 앞서 찾아진 알약 영역이 어떤 모양에 속하는지를 분류하기 위해 Zernike Moment[11]를 적용했다. Zernike Moment는 Edge와 Polar Coordinate를 사용하는 방법이나 Hu Moment를 사용하는 것에 비해 더 Noise에 안정적이다[25].

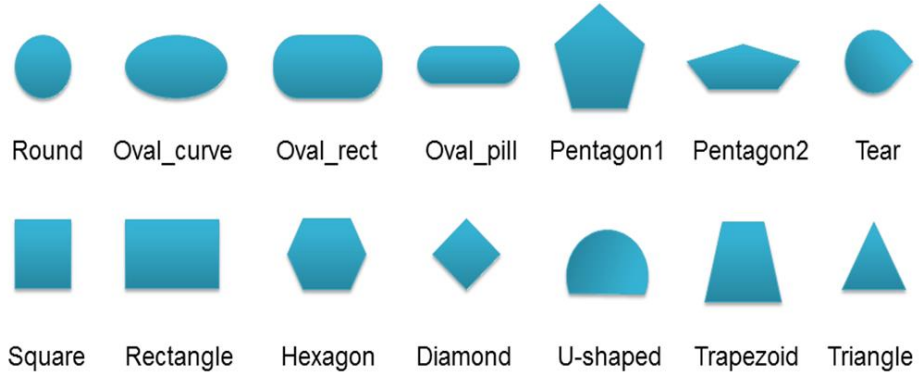


Figure 10. 분류에 사용된 14가지 모양 클래스

NLM 데이터베이스 중 Consumer Quality Image가 아닌 Reference Image로부터 각 클래스를 대표할 수 있는 Zernike Moment 값을 모멘트 차수 (0,0), (1,1), (2,0), (2,2), ..., (5,5) 총 12개로 표현하였다. 실제로 획득된 알약 영역과 Zernike Moment 값의 차이를 RMS로 계산하여 가장 근접한 모양을 택하였다. 이를 수식으로 표현하면 다음과 같다.

$$A_{mn} = \frac{m+1}{\pi} \iint_{x,y} I(x,y) [V_{mn}(x,y)]^* dx dy$$

$$V_{nm}(\rho, \theta) = R_{nm}(\rho) e^{jm\theta}$$

$$R_{nm}(\rho) = \begin{cases} \sum_{l=0}^{(n-m)/2} \frac{(-1)^l (n-1)!}{l! [\frac{1}{2}(n+m)-l]! [\frac{1}{2}(n-m)-l]!} \rho^{n-2l} & \text{for } n-m \text{ even} \\ 0 & \text{for } n-m \text{ odd} \end{cases}$$

여기서 m과 n값은 위에서 언급한 12개가 사용되었고 결과 값으로는 A값 12개를 사용하였다. 각 클래스 별로 오차는 RMS 값을 사용

하였으며 다음과 같이 계산한다.  $\tilde{A}_{mn}(C_i)$  를 각 클래스를 대표하는 Zernike Moment 값이라고 했을 때,

$$RMS(C_i) = \sqrt{\frac{\sum_m \sum_n \{A_{mn} - \tilde{A}_{mn}(C_i)\}^2}{12}}$$

RMS가 가장 작은 클래스 C가 알약에 해당하는 모양 클래스가 된다.

### 3.5 글자부 획득

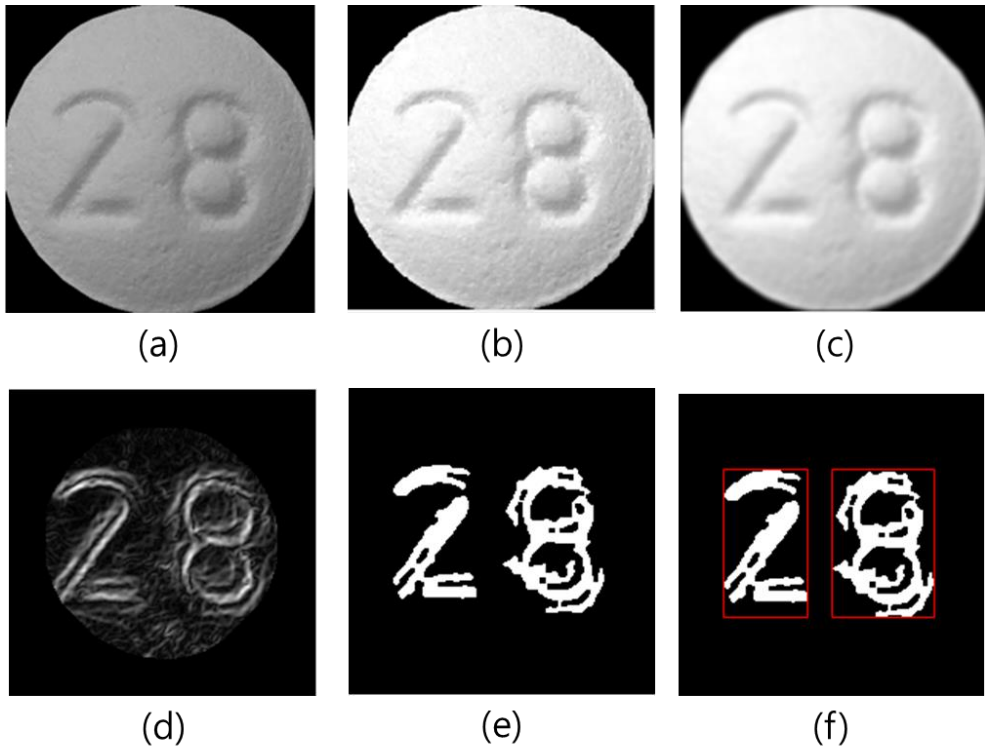


Figure 11. 글자부 강화 과정. (a) Gray-scale 이미지. (b) 밝기 조정. (c) Gaussian Filter. (d) Gradient Magnitude 영상. (e) 이진화. (f) Bounding Box 추출

알약에 새겨진 각인 글자는 보는 각도에 따라서 그리고 주변 빛의 환경 및 글자 획의 크기에 따라서 다양하게 변화할 수 있다. 따라서 이러한 변화에 견실하게 대응하기 위해 앞서 획득된 알약 영역 이미지에서 이미지의 밝기를 조정하여 다음 Gaussian Filter를 적용하였다. 앞서 획득된 알약 영역 이미지를 흑백으로 표현하면 Figure. 11(a)이며, 다양한 경우로 촬영되기 때문에, 글자 각인 영역의 밝기 차이가 뚜렷하지 않은 경우를 보완하기 위해, Figure. 11(b)와 같이 Intensity Adjustment를 수행한다. 이렇게 영상을 조정하면 글자 영역이 선명해지지만 부작용으로 알약 표면의 울퉁불퉁한 표면 또한 같이 강조될 수 있다. 따라서 이러한 일반적인 표면과 글자 영역을 구분해낼 필요가 있다. 일반적인 표면에서 발생하는 밝기 차이는 글자 영역에서 발생하는 밝기 차이의 영역에 비해 상대적으로 작기 때문에 Gaussian Filter를 이용하여 노이즈를 제거해주었다. 예를 들면, 알약의 배경 표면에서는 밝기 차이가 5x5픽셀 사이에서 크게 나타나는 반면, 글자 영역에서는 그 윈도우의 크기를 15x15이상 잡아야 밝기 차이를 구분할 수 있다.

각인 글자는 Figure. 11(c)에서 볼 수 있듯이, 획에 따라 어두워졌다가 밝아지는 특정한 패턴을 지니고 있다. 그림의 숫자 '2'의 경우, 글자 획의 윗부분은 그림자 효과에 의해 어두워지며 일정한 두께의 그림자 아래로는 깔여있는 표면에 빛이 닿음으로써 알약의 배경과 같은 색깔을 유지하게 된다. 글자 획의 아랫부분은 각도에 의해 주변보다 밝아진다. 이는 조명의 방향과 획의 방향에 따라서 유동적으로 변하게 된다. 따라서 방향과 획의 방향에 무관하면서 글자의 획을 뚜렷하게 보이

도록 강화하기 위하여 Gradient의 크기 값을 사용하였다. Gradient 방법을 사용하면 각인 글자처럼 배경과 글자의 색이 뚜렷이 구분되지 않더라도 Figure. 11(d)에서 볼 수 있듯이 충분히 경계를 추출해낼 수 있다.

그 뒤, 히스토그램에 기반한 Otsu Threshold를 적용하여 영상을 이진화 하고 노이즈들을 정리하면 Figure. 11(e)를 획득할 수 있다. 수식으로는 다음과 같이 표현된다.  $i$ 를 Intensity의 밝기 그리고  $h_i$ 를  $i$ 값의 히스토그램이라고 하면 전체 픽셀 수  $H$ 는

$$H = \sum_i h_i$$

이다. 여기서 상위 30%에 해당하는 픽셀 Threshold를 찾기 위해 다음 수식을 만족하는  $c$ 값 중 가장 큰 값을 찾는다.

$$\sum_{i=0}^c h_i < 0.7H$$

이렇게 찾아진  $c$ 를 바탕으로 Threshold  $t$ 는 다음과 같이 계산한다.

$$t = \frac{\sum_{i=c}^{255} ih_i}{\sum_{i=c}^{255} h_i}$$

여기서 Box를 찾기 전에 주변부에 노이즈가 생길 수 있기 때문에 노이즈를 제거해준다. Connected Component를 계산하고, 각각의 이진 영역에 Labeling을 한다. 그 뒤 각 영역별로 글자가 가져야할 특징을 이용해 글자와 글자가 아닌 영역을 걸러낼 수 있다.

- (1) Aspect Ratio < 3

$$(2) \quad 100 < \text{Area} < 20000$$

$$(3) \quad \text{Eccentricity} < 0.995$$

$$(4) \quad \text{Solidity} > 0.3$$

$$(5) \quad 0.2 < \text{Extent} < 0.9$$

여기에 수직 · 수평방향의 히스토그램 분석을 하면 글자를 감싸는 빨간색 상자를 계산할 수 있다.

### 3.6 글자 학습 및 인식

글자 인식은 일반적으로 OCR과 같은 도구를 사용하지만, OCR은 배경과 글자의 색이 흑백일 경우에만 가능하며 각인글자에서는 사용하기 힘들다. 물론 영상 처리를 통하여 Fig. 6(d)와 같이 흑백의 글자로 만들어 내긴 하였으나 현재 상용되는 OCR은 글자의 일부가 끊겨 있거나 내부의 잡음이 많은 경우 잘못 인식되는 경우가 많다. 따라서 본 논문에서 제안하는 시스템에서는 글자 학습을 따로 하여 각인 글자에 특화된 OCR을 만들었다.

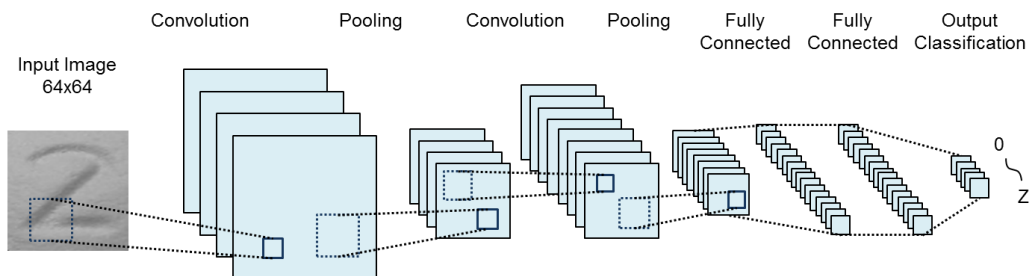


Figure 12. CNN Deep Learning의 원리

사용된 학습 방법은 Figure. 12에서 볼 수 있듯이 CNN Deep Learning이다. 이는 최근의 Natural Scene Text Recognition에서 많이 사용되는 방법으로써 다량의 이미지 데이터를 사용하고 인체의 뉴런과 유사한 모델로 구성함으로써 특정한 Feature를 설정해주지 않더라도 학습을 통해 이미지에서 필요한 Feature를 스스로 구성하게 된다. 이미지 데이터들을 얻기 위해 NLM 데이터베이스[8]의 Reference Image들에서 글자들을 수동으로 추출하고 합성 이미지들을 만들었다. 글자는 영문자 26자 숫자 10자로 총 36자이며, 각 글자 별 학습에 사용한 데이터 수는 수동 추출 이미지 2000장, 합성 이미지 1000장으로 총 3000장을 사용하였다.

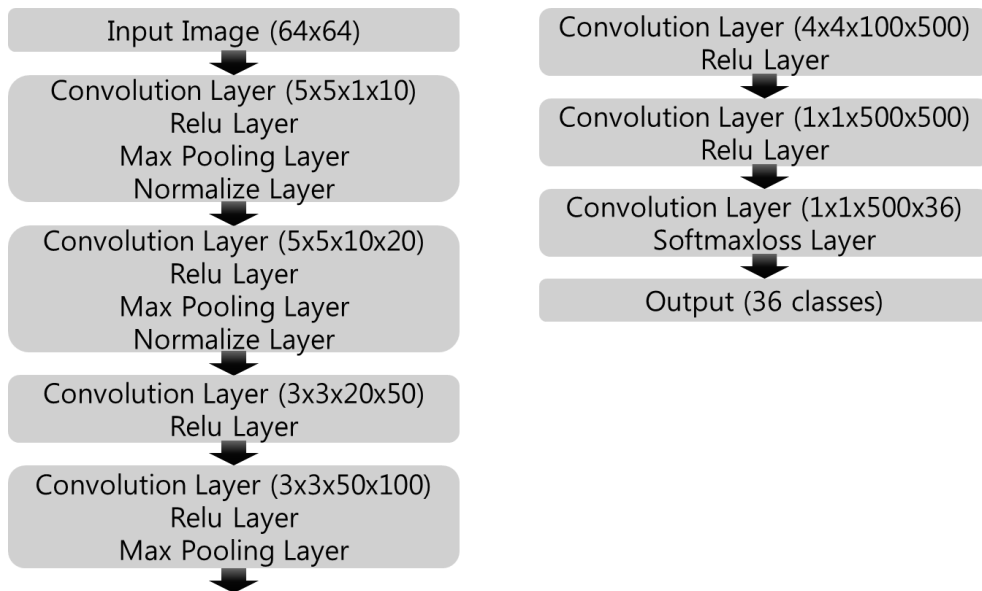


Figure 13. 각인 글자 인식을 위해 사용된 CNN 네트워크 구성

CNN의 네트워크 구성은 Figure 13과 같다. 먼저 가로 세로 64 픽셀 크기의 흑백 이미지를 Input으로 사용한다 그 뒤, 5x5 filter 10개를 Convolution 연산으로 계산한다. 그러면 차원은 60x60x10이 된다. 여기서 비선형 Layer인 Relu Layer와 Max Pooling Layer를 계산한다. Max Pooling Layer의 경우 윈도우의 크기를 2x2로 하였기 때문에 차원은 반으로 줄어 30x30x10이 된다. 그 뒤, 학습의 성능을 올려주기 위해 Normalize Layer를 통과한다 Normalize에 사용된 Parameter 값들은  $n=5$ ,  $kappa=1$ ,  $alpha=0.0001/5$ ,  $beta=0.75$ 이다.

그 뒤 다시 한번 5x5 filter를 거치면 차원은 26x26x20이 되고 앞서 구성한 것과 마찬가지로 같은 변수로 구성하면, 차원은 13x13x20이 된다. 그리고 3x3 Convolution Layer와 Relu Layer를 두 번 거치면 차원은 9x9x100이 되고, Max Pooling Layer를 계산하면 차원은 4x4x100이 된다. 마지막으로 Fully Connected Layer를 4x4x100x500, 1x1x500x500, 1x1x500x36으로 구성하였다. 마지막의 36개는 A부터 Z까지 그리고 0부터 9까지의 총 36개 글자 Class를 의미한다.

이렇게 학습된 OCR을 이용하여 앞서 3.5절에서 추출된 글자부박스를 원본 이미지에서 추출하고 글자를 인식 시킴으로써 글자 정보를 획득하게 된다.



## 제 4 장 실험 및 결과

### 4.1 Database

본 알고리즘을 평가하기 위해 대량의 알약을 실제로 촬영하는 것이 가장 좋지만, 일반적인 방법으로는 알약을 구하기가 매우 어렵다. 따라서, National Library of Medicine(NLM)에서 제공하는 데이터 베이스를 사용하였다. 이 데이터베이스는 Figure. 14의 좌측과 같이 일반 사용자 퀄리티의 이미지(Consumer Quality Image)도 있고, 우측과 같이 제한적인 환경에서 촬영된 이미지(Reference Image)도 있다. 각각 500장을 사용하였으며, Consumer Quality Image는 본 알고리즘의 평가를 위해, 그리고 Reference Image 500장은 다른 알고리즘에서 사용했었기 때문에 비교를 위해 사용하였다. Consumer Quality Image의 크기는 2048x1536에서부터 4416x3312까지 다양했다. 알고리즘은 MATLAB을 사용하여 작성되었다. 컴퓨터의 성능은 CPU Intel i7 950이다.



Figure 14. 이미지 데이터베이스 (NLM)

결과는 총 5가지로 분석하였다. 모양 정보는 알약의 분류에 상당한 비중을 차지하기 때문에 모양을 얼마나 제대로 분류 했는가는 중요한 지표이다. 또한 알약과 같이 각인 글자가 있는 영상에서 글자가 있는 영역을 찾아내는 능력은 글자 인식을 정확히 하기 위해 중요한 지표이다. 뿐만 아니라 찾아진 각인 글자 박스에서 정확한 글자로 인식해내는 지와 전체적인 글자 인식 성공률도 결과로 산출하였다. 마지막으로 모양과 글자를 조합한 알약 인식 성공률도 계산하였다.

## 4.2 Shape Matching Rate

이 실험에서 알약과 가장 가까운 모양의 5순위를 산출하였다. 그 결과는 Figure. 15와 같다.

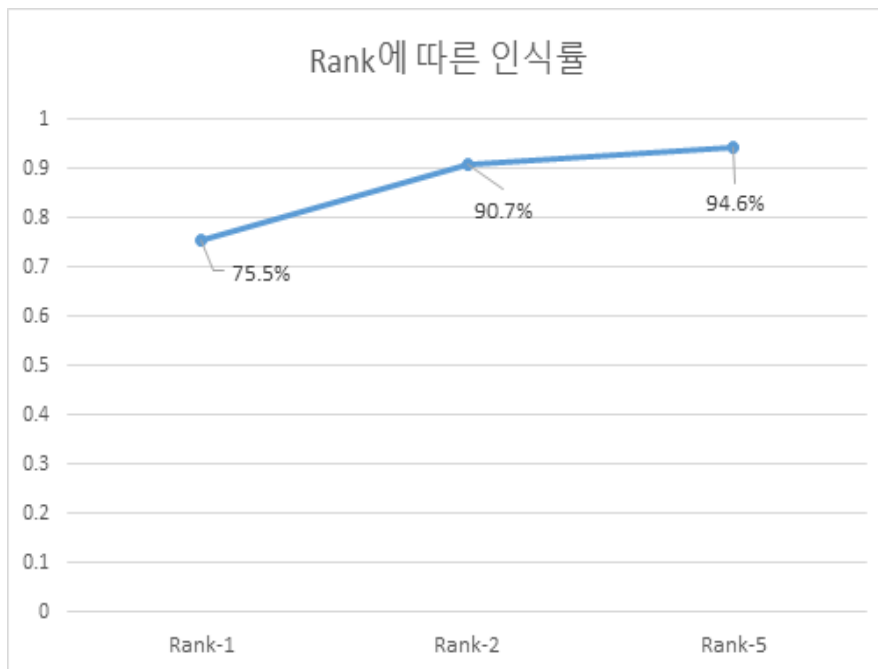


Figure 15. Rank에 따른 모양 인식률

Rank-1이란 모양을 인식하여 가장 가능성이 높은 모양을 추출했을 때 이것이 정답일 확률이며, Rank-2는 가장 가능성이 높은 모양 2순위안에 정답이 있을 확률이다. Rank-5는 마찬가지로 가능성 높은 모양 5순위안에 정답이 있을 확률이다.

Rank-1의 결과가 75.5%정도가 나왔는데, Rank-2에서는 결과가 90% 이상으로 좋아지는 것을 확인할 수 있다. 이는 모양 인식을 위해 알약 영역을 자르게 되는데 이 때 완벽하게 원형으로 잘리는 것이 아니다 보니 도형의 둘레에 각이 생기게 되고 Round형태와 Pentagon1 형태 사이에서 오차로 잘못 분류되었기 때문에, 실제로 Rank-5는 Rank-2에 비해 3개의 후보가 추가되었음에도 불구하고 Rank-1에서 Rank-2로 바뀔 때 인식률이 상승한 폭에 비해서는 아주 작다. 즉, 현재 알고리즘에서 성능을 더 올리기 위해서는 둘레에 각을 줄여주고 더 완벽하게 알약 영역을 잘라내야 한다.

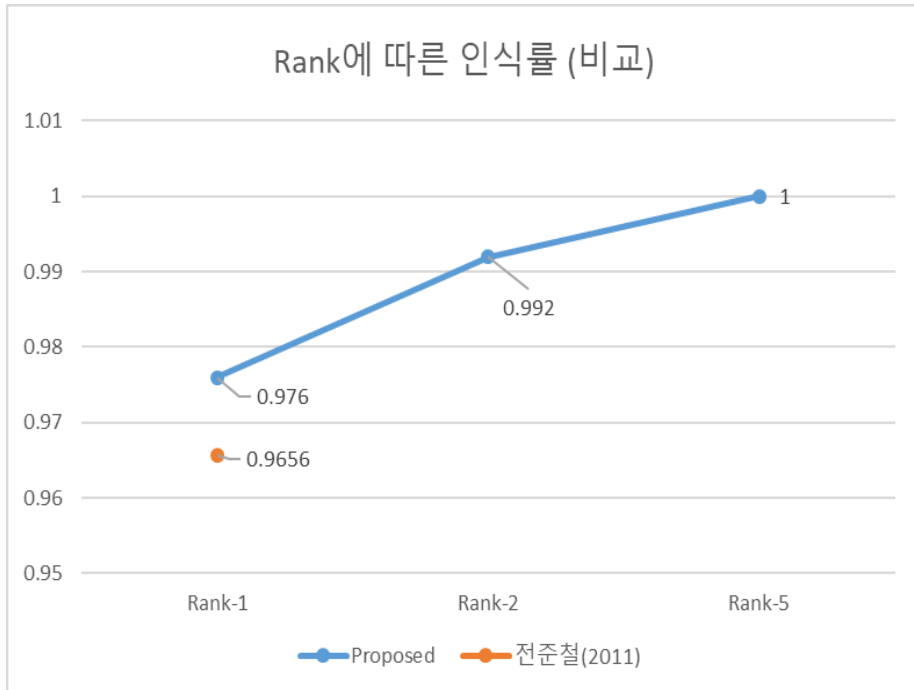


Figure 16. Rank에 따른 모양 인식률 (Reference 이미지를 이용한 비교)

Figure. 16은 비교를 위해 Reference Image로 테스트한 결과이다. 2011년 전준철[14]의 결과에서는 Rank-1밖에 없기 때문에 Rank-1만 비교하였으며, 본 연구에서 제시하는 방법의 인식률이 조금 더 높았다.

### 4.3 Text Box Detection Rate

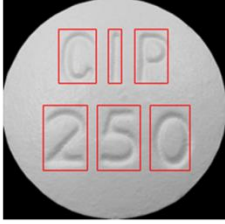


Example Image			
# of Characters	6	3	5
# of Correct Boxes	6	1	3
Detection Rate	100%	33.3%	60%

Figure 17. Text Box Detection Rate 계산 방식

Text Box Detection Rate는 정확하게 찾아진 Bounding Box의 수를 전체 글자 수로 나눈 값으로 정의하였다. 예를 들면, Figure. 17에서, 첫 번째 그림은 총 글자 6개에 제대로 찾아진 박스의 개수가 6개이므로 Detection Rate는 100%이다. 두 번째 그림은 총 3글자에 제대로 찾아진 박스는 1개이므로 Detection Rate는 33.3%이다. 세 번째 그림은 총 5글자에 3개가 찾아졌으므로 Detection Rate 60%이다. Rate의 평균 값은 Table 1과 같다.

Table 1. Text Box Detection Rate Result

	Proposed	Suntronsuk et al. (2016)	Kasar et al. (2007)
Text Box Detection Rate	<b>87.5%</b>	56.67%	21.21%

같은 이미지 데이터베이스를 사용하였으며 같은 방법으로 Rate를 계산한 Suntronsuk[23], Kasar et al.[26]의 연구 결과와 비교하였다. Suntronsuk의 방법에 비해 상당히 성능이 개선됨을 확인할 수 있었다. 이러한 차이가 발생하는 원인은 Suntronsuk의 접근 방법이 Edge와 이진화를 기반으로 하고 있기 때문이며, 빛 환경에 따라 Edge가 제대로 나오지 않는 경우에는 박스 검출에 실패하게 된다. 반면, 제안된 방법은 Gradient 방법에 기반하였기 때문에 빛 효과에 의한 실패율을 최소화 할 수 있었다.

## 4.4 Character Recognition Rate

Text Box Detection Rate와 별개로 알약에 존재하는 글자들을 얼마나 찾았는가, 그리고 얼마나 정확하게 찾아내는 지도 성능을 결정하는 중요한 요소이다. 따라서 이를 위해 3가지의 통계적 지표로 글자 인식률을 산출하였다.

- Retrieved item : 실제 찾아진 Bounding Box로부터 인식시켜 나온 글자
- Relevant item : 알약에 새겨진 실제 글자

$$precision = \frac{\{relevant\ items\} \cap \{retrived\ items\}}{\{retrived\ items\}}$$

$$recall = \frac{\{relevant\ items\} \cap \{retrived\ items\}}{\{relevant\ items\}}$$

$$F-score = \frac{2 \times precision \times recall}{precision + recall}$$

Table 2. Character Recognition Rate Result

	Proposed	Suntronsuk et al. (2016)	Kasar et al. (2007)
Precision	<b>0.776</b>	0.388	0.144
Recall	<b>0.835</b>	0.410	0.152
F-score	<b>0.786</b>	0.399	0.148

마찬가지로 Suntronsuk의 방법과 비교한 결과는 Table 2와 같다. 모든 면에서 제안된 방법이 높은 성능을 보였다. 이러한 이유는 Suntronsuk의 방법에서 영상 이진화 과정 중 local thresholding을 사용하게 되는데, 각인 글자는 배경과 색이 유사하기 때문에 노이즈가 발생할 확률이 높다. 반대로 제안된 방법은 Gaussian Filter와 Gradient Filter를 사용함으로써 이진화에서 발생하는 노이즈를 최소화 하였다.

## 4.5 Text Recognition Rate & Recognition Success Rate

글자들이 모여서 하나의 텍스트를 이루고 텍스트가 알약을 판별하는 구분 기호로 사용된다. 전체 글자 인식률 및 모양과 글자를 모두 포함한 인식 성공률은 Table 3과 같다.

Table 3. Text Recognition Rate & Recognition Success Rate Result

		Consumer Quality Image		Reference Image	
		Proposed	Suntronsuk et al. (2016)	Proposed	Jiye Yu et al. (2015)
Text Recognition Rate	Without parsing	55.4%	22.61%	73.2%	-
	With parsing	73%	-	92.8%	-
Recognition Success rate (with parsing)		58.4%	-	90.4%	90.46%

단순한 텍스트의 인식율은 55.4%이고 Parsing을 통해 데이터베이스에 존재하지 않는 텍스트는 가장 연관성이 높은 다른 텍스트로 검색하여 대체할 경우 73%이다. 나머지 27%의 오류는 대부분 글자가 ‘M’ , ‘b’ 와 같이 한 글자 일 때가 많았다. 제안된 시스템에서는 글자가 한 글자 일 경우 Text Box가 크게 잡히게 되고, 경우에 따라 이를



여러 개의 작은 단위로 나누게 되기 때문이다. 또 글자가 한 글자일 경우에는 파싱이 불가능한 문제도 있다. 따라서 글자 크기에 영향을 덜 받는 알고리즘을 구축한다면 인식률은 크게 개선될 수 있다. 글자와 모양을 모두 포함한 인식률은 58.4%이다. 실제 환경에서의 인식률은 Suntronsuk의 연구와 비교할 경우 압도적으로 높다. 본 알고리즘의 인식률이 높아지기 위해서는 글자 인식률을 더 높여야 한다. 글자 인식률이 개선되면 충분히 90%까지 개선될 수 있다. 또한, 글자 인식에 실패했을 경우, 사용자에게 재촬영을 요구한다면, 인식률은 더 높아질 수 있다.

Table 3 오른쪽의 Reference Image에 대한 결과는 빛이나 그림자 효과가 확실하게 제거되어 있는 데이터베이스에서 결과를 계산한 것이다. 기존의 논문과 차이가 거의 없었으나, 기존의 논문은 Descriptor를 사용한 방식이고 본 논문에서 제시하는 방법은 Text 인식을 사용한 방식이기 때문에 차후 상용화의 관점에서 본다면 본 논문의 방식이 더 우수하고, 더 광범위한 주변 환경을 커버할 수 있다.

## 제 5 장 결론

본 논문에서는 일반 환경에서 스마트폰으로 촬영된 알약 영상으로부터 글자와 모양을 인식하는 방법을 제안한다. 알약이 포함된 영상에서 알약 영역을 특정 짓기 위해 Saliency Map 사용하고 그림자 제거를 하였다. Zernike Moment를 이용하여 회전 불변성 모양 인식을 하였으며, 획득된 알약 영역에서 Gaussian 및 Gradient Filter를 이용해 기존보다 노이즈에 강인하게 글자부를 강화하였다. 알약의 글자를 감싸는 박스를 추출하고, CNN으로 학습된 각인 글자 인식기를 사용해 최종적인 글자 정보를 획득한다. 500개의 이미지에 대해 척박한 실제 사용환경에서도 58.4%의 인식률을 얻었다. 이는 주변 환경에 더 견실하면서 기존의 연구들을 상당히 뛰어넘는 결과이다. 모양 인식률과 글자 박스 추출 성능을 개선한다면 인식률은 더욱 개선될 것이며 일반 사용자들에게도 충분히 적용가능하다.

## 참고 문헌

- [1] '의약품사전', available at  
<http://terms.naver.com/medicineSearch.nhn?mode=nameSearch&query=>, accessed 17.05.23
- [2] '약학정보원', available at  
[http://www.health.kr/drug\\_info/sb/search.asp](http://www.health.kr/drug_info/sb/search.asp), accessed 17.05.22
- [3] 'WebMD pill identification tool', available at  
<http://www.webmd.com/pillidentification/>, accessed 17.05.22
- [4] 'RxList pill identification tool', available at  
<http://www.rxlist.com/pillidentification-tool/article.htm>, accessed 17.05.22
- [5] 'Healthline pill identifier', available at  
<http://www.healthline.com/pill-identifier>, accessed 17.05.23
- [6] 'DailyMed-national library of medicine', available at  
<http://dailymed.nlm.nih.gov/dailymed/index.cfm>, accessed 17.05.22
- [7] 'Drug.com Pill Identifier', available at  
[http://www.drugs.com/pill\\_identification.html](http://www.drugs.com/pill_identification.html), accessed 17.05.23
- [8] 'U.S National Library of Medicine', available at  
<http://www.nlm.nih.gov/>, accessed 17.05.20
- [9] Harel, Jonathan, Christof Koch, and Pietro Perona. "Graph-based visual saliency." NIPS. Vol. 1. No. 2. 2006.

- [10] J. Canny, "A computational approach to edge detection," IEEE Transactions on pattern analysis and machine intelligence, vol. 6, pp. 679–698.
- [11] Teague, Michael Reed. "Image analysis via the general theory of moments." JOSA 70.8 (1980): 920–930.
- [12] Hartl, Andreas, Clemens Arth, and Dieter Schmalstieg. "Instant Medical Pill Recognition on mobile phones." Proceedings of 16th IASTED international conference on robotics and applications (RA 2011). 2011.
- [13] Cunha, António, Telmo Adão, and Paula Trigueiros. "HelpmePills: a mobile pill recognition tool for elderly persons." Procedia Technology 16 (2014): 1523–1532.
- [14] 전준철, and 김동선. "모양분류와 컬러정보를 이용한 내용기반 약 영상검색 시스템☆." 인터넷정보학회논문지 12.6 (2011): 117–128.
- [15] Annasaro, Dr E., and A. Hema. "Color and shape feature extraction and matching in pill identification systems." International Journal of Computer Science and Information Technologies 5.2 (2014): 1011–1015.
- [16] Hu, Ming-Kuei. "Visual pattern recognition by moment invariants." IRE transactions on information theory 8.2 (1962): 179–187.
- [17] Lee, Young-Beom, Unsang Park, and Anil K. Jain. "Pill-ID:

Matching and retrieval of drug pill imprint images." Pattern Recognition (ICPR), 2010 20th International Conference on. IEEE, 2010.

[18] Lowe, David G. "Distinctive image features from scale-invariant keypoints." International journal of computer vision 60.2 (2004): 91–110.

[19] Ojala, Timo, Matti Pietikainen, and Topi Maenpää. "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns." IEEE Transactions on pattern analysis and machine intelligence 24.7 (2002): 971–987.

[20] Lee, Young-Beom, et al. "Pill-ID: Matching and retrieval of drug pill images." Pattern Recognition Letters 33.7 (2012): 904–910.

[21] Yu, Jiye, Zhiyuan Chen, and Sei-ichiro Kamata. "Pill recognition using imprint information by two-step sampling distance sets." Pattern Recognition (ICPR), 2014 22nd International Conference on. IEEE, 2014.

[22] Yu, Jiye, et al. "Accurate system for automatic pill recognition using imprint information." IET Image Processing 9.12 (2015): 1039–1047.

[23] Suntronsuk, Siroratt, and Sukanya Ratanotayanon. "Pill image binarization for detecting text imprints." Computer Science and

Software Engineering (JCSSE), 2016 13th International Joint Conference on. IEEE, 2016.

[24] Suntronsuk, Siroratt, and Sukanya Ratanotayanon. "Automatic text imprint analysis from pill images." Knowledge and Smart Technology (KST), 2017 9th International Conference on. IEEE, 2017.

[25] 김선중. "실루엣 영상기반 보행 분석을 위한 형태기술자의 성능 평가." 지능정보연구 15.4 (2009): 53–64.

[26] Kasar, Thotreingam, Jayant Kumar, and A. G. Ramakrishnan. "Font and background color independent text binarization." Second international workshop on camera-based document analysis and recognition. 2007.

# Abstract

## Shape and Text Imprint Recognition of Pill Image Taken with a Smartphone

Dae-wook Kim

Department of Mechanical and Aerospace Engineering

College of Engineering

The Graduate School

Seoul National University

Numerous pills are manufactured and sold to benefit people's lives, but these pills can cause fatal threats for people if misused or abused. Although pills can be searched through the internet and smart phones, users need to input shapes, colors, and letters. This reduces accessibility particularly to the elderly. Therefore, it is necessary to acquire information from the pill image with minimum efforts. In this paper, we propose a method to recognize characters and shapes in tablet images taken by smartphone. To specify the pill area, saliency map and shadow effect removal are adapted.

Through the Zernike moment, the shape information is obtained. The box surrounding the pill is extracted through Gaussian and gradient filter. The character information is determined by OCR learned through CNN. Recognition experiments are applied on NLM database and result shows 58.4% rank-1 matching accuracy when testing with 500 severe real environment image. This is a far superior result to the state-of-the-art ones, and if the bounding box detection is complemented a bit more, it can be applied to general users as well.

**Keywords** : Character Recognition, Deep Learning, Image Processing, Image Retrieval, Pill Identification, Zernike Moment

**Student Number** : 2014-21871